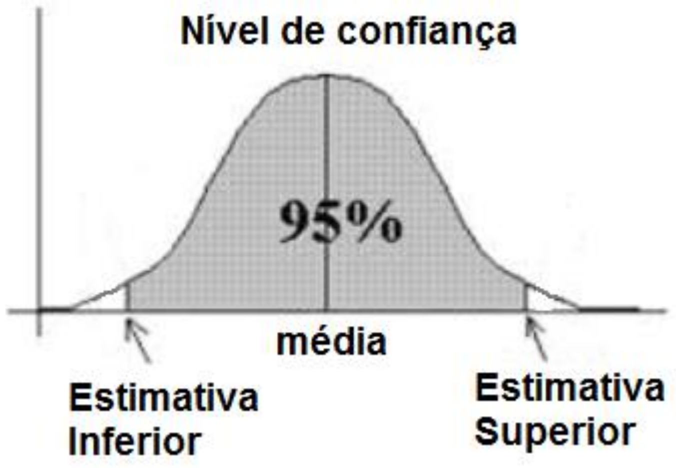
Inferência Estatística I

INTERVALOS DE CONFIANÇA

TUANY CASTRO

Intervalos de Confiança

- Os estimadores vistos até aqui eram estimadores pontuais, pois fornecem apenas um valor;
- Como os estimadores são variáveis aleatórias, podemos fornecer uma estimativa mais informativa que inclua uma medida de precisão;
- Esse método é denominado intervalo de confiança e é obtido a partir da distribuição amostral dos estimadores.



Intervalo de Confiança para Média

- O intervalo de confiança é formado por dois valores que garantem uma precisão γ na estimação do parâmetro;
- O intervalo de confiança para a média μ com coeficiente de confiança γ é dado por:

$$\left[\bar{X} - z * \frac{\sigma}{\sqrt{n}} ; \bar{X} + z * \frac{\sigma}{\sqrt{n}}\right]$$

Exemplo: Uma máquina enche pacotes de café com uma variância igual a $100g^2$. Ela estava regulada para encher pacotes com 500 g, em média. Agora, ela se desregulou, e queremos saber a nova média. Uma amostra de 25 pacotes apresentou uma média de 485g. Qual o intervalo de confiança com 95% de confiança para a média do peso dos pacotes?

Intervalo de Confiança para Média

Exemplo: Uma máquina enche pacotes de café com uma variância igual a $100g^2$. Ela estava regulada para encher pacotes com 500 g, em média. Agora, ela se desregulou, e queremos saber a nova média. Uma amostra de 25 pacotes apresentou uma média de 485g. Qual o intervalo de confiança com 95% de confiança para a média do peso dos pacotes?

Temos $\bar{X} = 485g$, $\sigma^2 = 100g^2$, $n = 25$. Para 95% de confiança, o valor z da distribuição Normal é 1,96. Portanto:

$$\left[\bar{X} - z * \frac{\sigma}{\sqrt{n}}; \bar{X} + z * \frac{\sigma}{\sqrt{n}}\right] = \left[485 - 1{,}96 * \frac{10}{\sqrt{25}}; 485 + 1{,}96 * \frac{10}{\sqrt{25}}\right] = [481{,}1; 488{,}9]$$

Interpretação: Com 95% de confiança, o intervalo de 481,1g a 488,9g contém o valor da verdadeira média de peso dos pacotes preenchidos por essa máquina. Ou seja, com 95% de confiança está abaixo dos 500g especificado.

Intervalo de Confiança para Média

$$\left[\bar{X} \pm z * \frac{\sigma}{\sqrt{n}}\right] = [LI, LS]$$

$$LS - LI: \text{Amplitude}$$

- Amplitude: tamanho do intervalo, diferença entre limite inferior e limite superior;
- No exemplo, a amplitude foi de 7,8;
- Quanto maior a confiança, maior a amplitude do intervalo, assim menor a precisão porém maior assertividade.
- Quanto menor a confiança, menor a amplitude do intervalo, assim maior precisão porém menor assertividade;

Intervalo de Confiança para Média

$$\left[\bar{X} \pm \underbrace{z * \frac{\sigma}{\sqrt{n}}}_{\text{Erro da estimação}}\right]$$

- Quanto maior a variância dos dados, maior o erro;
- Quanto maior o tamanho da amostra, menor o erro;
- No exemplo, o erro foi de 3,8.

Exercícios

Exercício 1: Por analogia a produtos similares, o tempo de reação de um novo medicamento pode ser considerado como tendo distribuição Normal com desvio-padrão igual a 2 minutos (a média é desconhecida). Vinte pacientes foram sorteados, receberam o medicamente e tiveram seu tempo de reação anotado. Os dados foram os seguintes (em minutos): 2.9, 3.4, 3.5, 4.1, 4.6, 4.7, 4.5, 3.8, 5.3, 4.9, 4.8, 5.7, 5.8, 5.0, 3.4, 5.9, 6.3, 4.6, 5.5, e 6.2. Obtenha um intervalo de confiança para o tempo médio de reação com 95% de confiança (z=1,96).

Exercício 2: Uma amostra de 25 observações de uma distribuição com média desconhecida e variância 16 foi coletada e forneceu uma média amostral de 8. Construa intervalos com confiança 90%, 95% e 99%. Comente as diferenças encontradas.

Exercícios

Exercício 3: Num grupo de pacientes, o nível de colesterol é uma variável aleatória com distribuição Normal de média desconhecida e variância 64 (mg/ml^2).

A) Para uma amostra de 46 indivíduos que forneceu nível médio de colesterol de 120 mg/ml, construa o intervalo de confiança de 95%.

B) Se você desejasse diminuir a amplitude do intervalo, quais seriam suas alternativas?

Exercício 4: O consumo de combustível é uma variável aleatória com parâmetros dependendo do tipo de veículo. Suponha que, para um certo automóvel, o desvio-padrão do consumo seja conhecido e igual a 2 km/l. Porém precisamos de informações sobre o consumo médio. Para tal, coletamos uma amostra de 40 automóveis desse modelo e observamos o seu consumo.

A) Quem seria um estimador do consumo médio para todos os automóveis desse tipo?

B) Se a amostra forneceu um consumo médio de 9,3 km/l, construa um intervalo de confiança de 90% para a média de consumo desses carros.

C) Qual deve ser o tamanho da amostra se eu quiser um erro de no máximo 0,3 com 95% de confiança?

Intervalos de confiança para Média com Variância Desconhecida

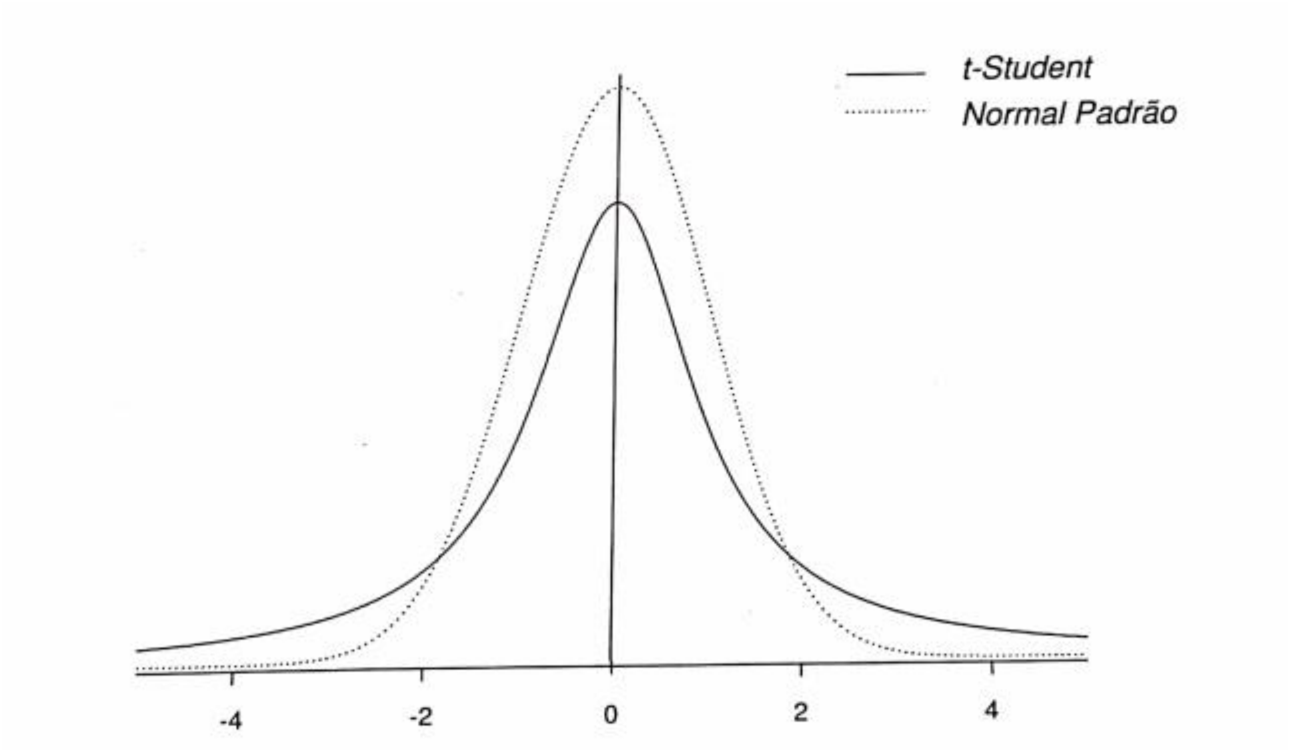
- Se o desvio-padrão é desconhecido, ele precisa ser estimado.
- Vimos que o melhor estimador para a variância é $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$.
- Temos que:

$$T = \frac{\bar{X} - \mu}{\sqrt{S^2/_n}} = \frac{\bar{X} - \mu}{S/_{\sqrt{n}}} \sim t_{n-1}$$

em que t_{n-1} corresponde à distribuição t-student com n-1 graus de liberdade. A t-student também tem densidade em forma de sino, entretanto as caudas tem mais massa do que a Normal(0,1).

A distribuição t-student se aproxima da distribuição normal conforme aumenta o tamanho da amostra. Para graus superiores a 120, voltamos a usar a distribuição normal.

Intervalo de confiança para Média com Variância Desconhecida



Intervalos de confiança para Média com Variância Desconhecida

- Se a variância σ^2 não for conhecida, o intervalo de confiança para a média toma a forma:

$$\left[\bar{X} - t * \frac{s}{\sqrt{n}} ; \bar{X} + t * \frac{s}{\sqrt{n}}\right],$$

em que $S^2 = \frac{1}{n-1}\sum_{i=1}^{n}(X_i - \bar{X})^2$ é a variância amostral e s o desvio-padrão (chamado erro-padrão); t é o quantil obtido a partir da distribuição t-student.

Exercícios

Exercício 5: Considere o exercício 1. Refaça o intervalo de confiança sem o conhecimento do desvio-padrão. Compare os intervalos.

Exercício 6: Admitindo que a pressão arterial em homens siga o modelo Normal, 7 pacientes foram sorteados e tiveram sua pressão medida com os seguintes resultados: 84, 81, 77, 85, 69, 80 e 79. Determine o intervalo de confiança para μ com coeficiente de confiança $\gamma = 95\%$.

Intervalo de Confiança para Proporção

- O intervalo de confiança para uma proporção p com coeficiente de confiança γ é dado por:

$$\left[\hat{p} - z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\,; \hat{p} + z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right]$$

em que $\hat{p}$ é a proporção observada na amostra.

Exemplo: Pretende-se estimar a proporção de cura, através do uso de um certo medicamento em doentes contaminados com cercaria, que é uma das formas do verme da esquistossomose. Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados. Que podemos dizer da proporção amostral na população geral com 95% de confiança?

Intervalos de Confiança

Exemplo: Pretende-se estimar a proporção de cura, através do uso de um certo medicamento em doentes contaminados com cercaria, que é uma das formas do verme da esquistossomose. Um experimento consistiu em aplicar o medicamento em 200 pacientes, escolhidos ao acaso, e observar que 160 deles foram curados. Que podemos dizer da proporção amostral na população geral com 95% de confiança?

Temos $\hat{p} = \frac{160}{200} = 0{,}8 \text{ e } n = 200$. Para 95% de confiança, o valor z da distribuição Normal é 1,96. Portanto:

$$\left[\hat{p} - z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}; \hat{p} + z * \sqrt{\frac{\hat{p}(1-\hat{p})}{n}}\right] = \left[0{,}8 - 1{,}96 * \sqrt{\frac{0{,}8 * 0{,}2}{200}}; 0{,}8 + 1{,}96 * \sqrt{\frac{0{,}8 * 0{,}2}{200}}\right] = [0{,}745; 0{,}855]$$

Interpretação: Com 95% de confiança, o intervalo de 0,745 a 0,855 contém o valor da verdadeira proporção de cura do medicamento.

Exercícios

Exercício 7: Numa pesquisa com 50 eleitores, o candidato José obteve 0,34 da preferência dos eleitores. Construa, para a confiança de 95%, um intervalo para a proporção de votos a serem recebidos pelo candidato supondo que a eleição fosse feita nesse momento.

Exercício 8: Uma amostra aleatória de 625 donas de casa revela que 70% delas preferem a marca de detergente A. Construa um intervalo de confiança para a proporção de donas de casa que preferem A com coeficiente de confiança de 90%.

Exercício 9: Numa pesquisa de mercado, desejamos estimar a proporção de pessoas que compram um determinado sabonete.

A) Se numa amostra de tamanho 100, a aceitação do produto foi de 0,8, qual o intervalo de confiança com 95% de confiança para a proporção de pessoas que compram o produto?

B) Se sabemos que a aceitação do produto é no mínimo 0,8, qual deve ser o tamanho da amostra para garantir um erro de 0,05 com coeficiente de confiança 90%?

Dimensionamento de amostras

- Vimos que o erro ao se construir um intervalo de confiança para uma média é da forma:

$$z * \frac{\sigma}{\sqrt{n}} \text{ ou } t * \frac{s}{\sqrt{n}}$$

- Se fixarmos um valor máximo para o erro, um coeficiente de confiança e tendo o valor de σ ou s, podemos obter o tamanho mínimo da amostra:

$$z * \frac{\sigma}{\sqrt{n}} \leq \text{erro ou } t * \frac{s}{\sqrt{n}} \leq \text{erro, então}$$

$$n \geq \left(z * \frac{\sigma}{erro}\right)^2 \text{ ou } n \geq \left(t * \frac{s}{erro}\right)^2$$